Postprint of TrAC Trends in Analytical Chemistry Volume 129, August 2020, 115913
DOI: https://doi.org/10.1016/j.trac.2020.115913

Performance parameters for analytical method validation: Controversies and discrepancies among numerous guidelines

Francisco Raposo[1]*; Carolina Ibelli-Bianco[2]
[1] Instituto de la Grasa, Consejo Superior de Investigaciones Científicas (IG-CSIC)
Campus Universitario Pablo de Olavide, Carretera de Utrera km 1, Edificio 46
41013 Seville, Spain
[2] Department of Hydraulics and Sanitation, School of Engineering of São Carlos, University of São Paulo. Avenida Trabalhador Sãocarlense, 400, Centro – São Carlos - SP - CEP: 13566-590 - Brazil
* To whom correspondence should be addressed:
(Tel.: +34 954611550; E-mail: fraposo@cica.es)

ABSTRACT

The main objective of method validation process is to prove that an analytical method is acceptable for its intended purpose. The necessity for laboratories to use fully validated methods is now universally accepted as a way to obtain reliable results. There are diverse documents for method validation including information about different performance parameters. The classical performance characteristics are accuracy, limit of detection, precision, recovery, robustness, ruggedness, selectivity, specificity and trueness. Unfortunately, contradictory information is normally present among the method validation documents used by laboratories. The inconsistency about the performance parameters can generate some degree of confusion in the complete method validation process. This manuscript addresses controversial and discrepant information, focusing specifically on several national and international method validation guidelines published by prominent organizations and institutions which serve as guidance to validate new analytical methods by practitioners working in different fields.

KEYWORDS

Analytical method validation; Accuracy; limit of detection; precision; recovery; robustness; selectivity; trueness.

ABBREVIATIONS

AAS - Atomic Absorption Spectroscopy; **AMV** - Analytical Method Validation; **ANOVA** - Analysis Of Variance; **C&D** - Controversies and Discrepancies; **CCα** - Decision limit; **CCβ** - Detection capability; **CDER** - Center for Drug Evaluation and Research; **CE** - Capillary Electrophoresis; **CV** - Coefficient of Variation; **DL** - Detection limit; **ECD** - Electron Capture Detector; **EMA** - European Medicines Agency; **GC** - Gas Chromatography; **GC-MS** - Gas chromatography–Mass Spectrometry; **HPLC** - High Performance Liquid Chromatography; **HG** - High Resolution; **ICP** - Inductively Coupled Plasma; **IR** - Infrared; **ISO** - International Organization for Standardization; **IUPAC** - International Union of Pure and Applied Chemistry; **LC-MS** - Liquid Chromatography Mass Spectrometry; **LOD** - Limit Of Detection; **LOQ** - Limit Of Quantitation; **MS** - Mass Spectrometry; **MV** - Method Validation; **NMKL** - Nordic Committee on Food Analysis; **NMR** - Nuclear Magnetic Resonance; **OLS** - Ordinary Least Squares; **r** - Correlation Coefficient; **R^2** -

Determination coefficient; **RE** - Relative Error; **RSD** - Relative Standard Deviation; **s or SD** - Absolute Standard Deviation; $\mathbf{s^2}$ - Variance; **TLC** - Thin-Layer Chromatography; **USFDA** - The United States Food and Drug Administration; **VAR** - Various; **WLS** - Weighted Least Squares; **α** - False positives; **β** - False negatives

CONTENTS

1. Introduction

Method validation (MV) is the process of proving that an analytical method is acceptable for its intended purpose. That means the ultimate objective of the MV process is to provide evidence that the method can provide reliable results. Analytical MV is carried out to ensure that every future measurement in routine analysis will be close enough to the unknown true value for the content of the analyte in the sample. It is absolutely important not to mix the terms analytical and bioanalytical methods as they both serve different purposes and cover different parameters for their particular validation procedures. Unfortunately, there is some misleading information in the literature because the term bioanalytical method validation is used to refer to the quantitative determination of drugs and/or metabolites in fluids and other biological matrices (blood, serum, plasma, urine, faeces, tissue skin). But really, this type of laboratory analysis that use such matrices should also be considered as analytical determinations. Thus, there are few techniques such as conventional chromatographic based methods (GC and HPLC) sometimes in combination with mass spectrometry (GC-MS and LC-MS) that can be used for diverse matrices. These techniques are very popular in routine laboratories belonging to different analytical environments. At this point it is appropriate to clarify that this document is focused on analytical MV and, therefore, bioanalytical chemistry and genuine biochemical analysis are outside its scope.

When a laboratory is interested in performing a new analytical procedure, one of the most important steps is its validation. The necessity for laboratories to use a fully validated method of analysis is now universally accepted and/or required within many sectors of analysis. In any case, although MV is an important requirement in the practice of chemical analysis, the general understanding among practitioners to why, when and what should be done for MV appears to be poor. This fact is due to frequent discrepancies among documents relating to MV published in the literature. As a consequence, there are some risks and problems when trying to work in the laboratory using contradictory definitions and requirements for the different validation parameters [1–6].

This manuscript has three main objectives. Firstly, to highlight the importance of the MV, drawing attention to the many problems that may be caused if an incorrect validation procedure is used. Secondly, to compile the numerous national and international regulatory documents or guidelines for analytical MV. Thirdly, to present a critical discussion among existing MV guidelines to emphasize possible pitfalls and expected trends that arise from MV to results assessment. Thus, important information including controversies and discrepancies (C&D) may be used as guidance by practitioners or scientists needing to validate new analytical methods.

2. Guidelines for MV

Many international guidelines and publications concerning MV were published in the literature. For this manuscript, the 37 different documents summarized in **Table 1** were evaluated [7–42]. The criteria for inclusion of guidelines was to try to compile the maximum number of documents previously reported in the literature. Previous comparative studies of MV guidelines used a limited number of documents, among 3-6 [1-6]. The documents can be classified according to diverse factors such as: i) matrix of samples (analytical versus biological); ii) national or international level; iii) area or discipline; iv) analytical technique; v) compounds analysed. In general, there are few MV guidelines dedicated to evaluate biological samples. Most of the documents are promoted by international organizations and regulatory agencies. The most frequent disciplines are pharmaceutical, environmental, toxicological and food analysis. The majority of documents can be used for any analytical

technique, although some of the documents were specific for chromatography determinations. Similarly, most of the documents were not focused to determine specific compounds but some of them are dedicated to pesticides analysis.

3. Inconsistencies among MV guidelines

3.1. Description of general factors

The realization of MV is not a single and universal procedure. The variability among MV guidelines may be related to the following different factors:

1st. Area of application and terminology. The biggest problem encountered about MV is the terminology employed in the extensive literature. When comparing documents, identical terms may be defined in different ways. In addition, some of the performance parameters are often used interchangeably and/or incorrectly. One of the reasons could be that the technical terms used for analytical methods vary in different sectors of analytical measurement. This ambiguity or misinterpretation in the terminology can lead in some instances to wrong scientific conclusions. It is important to consider that the harmonization in MV vocabulary is required for a discussion between scientists of the same or different analytical fields. For this purpose, the international vocabulary of metrology (VIM) was developed to describe measurements that can be used in different fields [43].

2nd. Particular purpose. Initially, analytical methods can be used for qualitative and quantitative determinations, although this document is only dedicated to the latter.

Furthermore, quantitative analytical methods can be used for different purposes, such as product development, process control, quality control and research. This fact can vary the MV procedure. For example, research validation works are normally carried out in perfect experimental conditions while the use of the same method in a routine laboratory needs a more systematic scheme for the internal validation procedure. Additionally, to check that method performance parameters are effective when the method is in repetitive use, validation should be appropriately evaluated in the laboratory including internal quality control activities.

3rd. Analytical techniques. There are different techniques to be used such as chromatography (GC, HPLC, TLC), capillary electrophoresis (CE), spectrophotometry (UV/VIS, IR, fluorescence, AAS, ICP) or spectrometric techniques (NMR, MS) as well as the hyphenated methods. They have their own special features which should be considered in detail for MV procedure.

4th. Validation parameters. The classical performance parameters are accuracy, precision, linearity and application range, limit of detection (LOD), limit of quantitation (LOQ), selectivity/specificity, recovery and robustness/ruggedness. It is possible that some validation documents consider complementary performance parameters such as carry-over, stability and system suitability studies.

5th. Experimental procedures. Although there is a general agreement among literature in terms of validation parameters, significant diversity exists with respect to the methodology employed. Many documents are usually restricted to general concepts [44] and there is frequently a lack of advice for the practical execution of MV studies [45]. Additionally, there are no official guidelines on the correct sequence of validation experiments, and the optimal sequence may depend on the method itself [46].

6th. Acceptance criteria. Only few criteria are normally provided to define the acceptance during MV. In part, this may be because acceptability is determined by the purpose served by the method and thus a broad overview of validation cannot address the differing requirements of each particular area of analysis.

3.2. Description of inconsistencies in performance parameters

3.2.1. Selectivity/Specificity

Obtaining a signal free unequivocally from the influence of other species contained in the sample is for reliable chemical measurement processes. In fact, the inexistence of interferences can be considered as the hallmark of any determination at laboratory level. Thus, if the analytical method is not free from the effect of possible interferences, all other performance parameters are less reliable [47].

Selectivity can be based on the detection system (e.g. atomic emission spectrometry) or separation process (e.g. chromatography). Hyphenated techniques (e.g. GC/LC-MS) can be applied when the demands for response signal free of interferences are especially high by combining selectivity from separation and detection processes.

[C&D-N1]. Terminology. The degree of interferences for analytical methods can be considered controversial because two terms such as selectivity and specificity co-exist. Despite the clear difference between the two terms, they are used interchangeably or erroneously, especially in the field of chromatography [48]. By one hand, the term specificity is used for single component analysis when a method is free from interferences and only determines the intended analyte. Thus, only a small number of biochemical methods relating to enzymatic and immunochemical determinations can be considered specific in the sense defined above. On the other hand, selectivity refers to multicomponent analysis as the extent to which it can determine one particular analyte or analytes in a complex mixture without interference from other components also present in the mixture. Additionally, IUPAC suggests that the term specific, in the analytical field, is considered as the ultimate of selectivity [49]. Also it is important to note the distinction of concepts included in the SANTE guideline for both parameters [33]. Selectivity is used to discriminate between the analyte of interest and other compounds while specificity is defined as the ability to provide signals to effectively identify the analyte. Therefore, this guideline differentiates among methodologies as selective/non-specific (e.g. GC-ECD), non-selective/specific (e.g. GC-MS) and selective/specific (e.g. HR-MS).

3.2.2. Calibration curve/Linearity/Response function

The analytical calibration represents the relationship between known amounts of the analyte in the sample and the response of the instrument. This procedure should be done during the early stage of the MV. Unfortunately, the experimental design for analytical calibration is not well described in all the documents. A detailed discussion on the strategy to carry out a calibration curve is beyond the scope of this article. Thus, the most important aspects in the experimental planning for analytical calibration are only cited: i) The type of the calibration samples, either matrix-containing or matrix-free; ii) The calibration methodology (external standard, internal standard or standard addition); iii) The range of concentrations and the distribution of the points along the calibration curve; iv) The number of replicate measurements for each calibration level; vi) The number of series or different calibration curves.

[C&D-N2]. Terminology. Many MV guidelines explaining that analytical calibration model should be chosen based on the linearity of experiments. Although the term linearity is generally accepted, this is not a very clear terminology [50].

[C&D-N3]. **Selection of the calibration model**. It must be pointed that the choice of an appropriate calibration model or response function is crucial for the quality of data that can be obtained with a given method during its routine application. In general, MV guidelines recommend to apply the simplest model that adequately describes the concentration–

signal relationship and the use of more complex models should be justified. However, this is not always easy to implement in practice due to two important subjects such as:

- The linearity of experiments. Although a linear relationship between instrument signal and analyte concentration is the simplest situation, the trends including non-linear response are very frequent for routinely laboratory work. Therefore, the use of quadratic or superior regression models may be necessary to avoid leverage points and deviations at low concentration levels [51].

- The selection of the fitting technique: Ordinary (OLS) versus weighted least squares (WLS). Calibration curves must be calculated by OLS linear regression, which assumes that variance is independent of the analyte concentration (homoscedasticity). But if the variance of the replicates at each concentration level is not constant through the linear range (heteroscedasticity), then a better option is to use the WLS regression method, which takes into account the individual variance values in each calibration point. Calibration ranges that span at least two or three orders of magnitude are usually related with significant heteroscedasticity, which is the very frequent situation for bioanalytical methods [52].

[C&D-N4]. Acceptance criteria. Different procedures were reported to evaluate the choice of the curve fitting such as graphically (scatter, residuals and sensitivity plots), statistically (ANOVA-lack of fit, Mandel test and significance of quadratic term test) and by numerical parameters (*r* and/or R^2, and % relative error or deviation from nominal values) [53]. One big problem is the lack of equivalence among some of the procedures typically applied to evaluate curve fitting [51]. In addition, one of the most controversial subjects relating to the evaluation of curve fitting is to check the linearity of a calibration curve by inspection of the correlation coefficient [50, 53]. At this point, it is important to clarify the difference between correlation and regression terms because many times they are used interchangeably. Correlation coefficient (*r*) describes the presence of a linear relationship between two observed variables, and the degree of association should be negative or positive. Contrarily, determination coefficient (R^2) does not care about the sign of the variation and it shows the association type by explaining the model. Therefore, *r* should be used to indicate the strength and direction of a linear relationship, while R^2 should be used to design the proportion of explained variance. However, although *r* and R^2 are widely reported for calibration curves, it is important to note that both parameters are unsuitable for goodness-of-fit regression evaluation [53]. In any case, the final decision about curve fitting should be made according the percentage of relative error (% RE) [51].

3.2.3. Accuracy

It is important to point out that accuracy is the most crucial parameter that any analytical method should address because it allows for estimating total error affecting the method [54].

[C&D-N5]. Terminology: one versus two parameters. In a strict sense, accuracy is only related to systematic error. This simple definition of accuracy as one simple parameter is thoroughly accepted in the bioanalytical field [55]. On the contrary, in a widespread sense, the term accuracy is considered as a function of random and systematic errors. Thus, accuracy is a dual parameter concept as a way to define the total analytical error. Then, the term precision is related to random error and the term trueness is related to systematic error [54]. There is an important difference between both precision and trueness. Although

the precision can be decreased, it cannot be fully eliminated. In contrast, trueness correction is in principle possible, although this is another controversial subject [19].

[C&D-N6]. Experimental procedure: single versus combined experiments. The evaluation of accuracy (or trueness) can be found together with precision in the form of combined experiments. On the contrary to parallel experiments, accuracy (or trueness) and precision are also determined by using separate tests. In this situation, precision of experiments should be checked previously to accuracy (or trueness) because precision affects evaluation of systematic error, but not vice versa. In any case, the accuracy samples should ideally be obtained from an independent source rather than the same from calibration curve and they should be as closely related to the unknown samples as possible.

3.2.4. Precision

Precision characterizes the closeness of agreement between the measured values obtained by replicate measurements on the same or similar objects under specified conditions [48]. Precision is generally assessed by repeated analysis of validation samples and it is usually expressed in the form of "imprecision" such as absolute standard deviation (*s or SD*), relative standard deviation (RSD), coefficient of variation (CV) or variance (s^2). Although the precision of an assay is constant over most of the range of an assay, the analysts should take into consideration that experimental precision shows a large variability, mainly decreasing at the extreme levels [56]. Therefore, testing precision is also essential at the bottom and top of the experimental range.

[C&D-N7]. Precision levels. Different terms are normally associated with random errors such as repeatability, intermediate precision and reproducibility [50]. The differences among precision levels are made by the concept of series or runs. Diverse factors such as operators, reagents, days and/or equipment can be varied during series/runs. The selection of the factors should be done according to the experimental conditions that will be found during the routine use of the analytical procedure.

On the other hand, it is important to note that the first type of precision that should be considered for MV is the instrument precision [57], also named as injection repeatability [3]. This instrument precision should be checked through replicate injections performed in repeatability conditions of the same solution at one considerable high concentration from the working range. It is calculated according to instrument signal, which depends on the technique used (e.g. Chromatography, checking the retention time and peak area; e.g. Ultraviolet and Visible measurements, checking the absorbance or transmittance at the selected wavelength).

[C&D-N8]. Terminology. Common terms to express the repeatability are within/intra-day, -assay, -batch and -run. Similarly, expressions for reproducibility of the analytical method are between/inter-day, -assay, -batch and -run. However, the expressions intra/within-day and inter/between-day precision are not preferred, because a set of measurements could take longer than one day or multiple sets could be analysed within the same day.

Another important subject about terminology is to distinguish between the terms intermediate precision and reproducibility because in some documents both terms are used interchangeably. The term intermediate precision should be used for single laboratory, while reproducibility should be associated with the random error obtained by many laboratories. Therefore, it should be pointed out that it is wrong to report the reproducibility precision for single laboratory and such a term should never be used. If the term reproducibility is used for one laboratory, to avoid misunderstanding, the term intra-

laboratory also must be used together. In this line, some documents can describe the reproducibility precision using two terms, intra-laboratory for single laboratory and inter-laboratory when multiple laboratories are validating one shared method.

3.2.5. Trueness

Trueness relates to the systematic error of a measurement system. Rigorously defined, refers to the agreement between the average of infinite number of replicate measured values and the true value of the measured quantity. In practice, trueness is evaluated from a finite but reasonably large number of measurements and reference values are used instead of the true value [54]. Trueness can be determined in one of four ways: i) By analysing a sample of known concentration (Certified Reference Material) similar to the routine sample and comparing the measured value to the true value; ii) Comparing test results from the method with results from an existing alternate method that is known to be reliable; iii) Based on the spiking of known amounts of analyte into sample matrix; iv) Using the technique of standard addition, which can be used in the case of matrix effect. The pros and cons of common approaches for determining trueness can be found elsewhere [58].

[C&D-N9]. Terminology. The trueness of an analytical method can be quantitatively expressed using three different terms such as bias, relative bias and recovery [59]. Firstly, bias is defined, in practice, as the difference between the mean obtained with a large number of replicate measurements and a reference value. Secondly, relative bias is calculated in similar manner considering the difference but also the reference value. Finally, recovery term should be used to denote the ratio of the concentration found versus the reference value. Therefore, the term trueness should be well explained in the validation document because frequently it is interchanged with other terms such as accuracy, bias and recovery.

3.2.6. Recovery

Although it is desirable to attain a recovery factor as close to 100% as possible, there is not a minimum established value. Therefore, an analytical method with low recovery could be suitable for a certain analyte if the sensitivity of the method is appropriate.

[C&D-N10]. Terminology. The general term recovery has been used in the literature in different situations. IUPAC explain that the term recovery is used in two distinct contexts that should be distinguished theoretically and also with a clear and different terminology [60]. By this way, the yield of a pre-concentration or extraction stage of an analytical process has been defined as absolute recovery, recovery factor or simply recovery. On the contrary, the ratio of observed value versus a reference value obtained using an analytical procedure that involves a calibration graph has been defined as relative or apparent recovery.

3.2.7. Limit of detection

This is an important figure of merit in the analytical chemistry field although it is not necessary to calculate during the process of validation of all analytical methods. The estimation of this parameter is especially important when trace and ultra-trace quantities of analyte are to be distinguished. Contrarily, LOD estimation for quantitative determinations at high concentration levels are omitted in the majority of MV guidelines. This is a greatly controversial performance parameter from both theoretical and experimental point of views with a lack of overall understanding and major differences in the terminology and the method of calculation.

[C&D-N11]. Terminology. In general, there are many options in the literature to describe measurement limits. The most frequent terms suggested by the chemical community to describe detection and quantification capabilities are critical value or decision limit; minimum detectable value or detection limit and minimum quantifiable value or quantification limit [61]. Some MV guidelines have presented alternative names but with similar definition. It is important to highlight that LOD is not the analyte level for deciding between detected and not detected [62]. The majority of definitions include terms such as confidence, probability and reliability, that denotes the use of statistics to calculate them. In fact, LOD is derived from the theory of hypothesis testing and the probabilities of false positives (α) and false negatives (β). Some of the conceptual problems caused by common definitions are solved by the use of alternative terms $CC\alpha$ (decision limit) and $CC\beta$ (detection capability) [63]. In addition, it is possible to find information about instrument LOD and method LOD. These terms refer to the instrument capabilities and the whole method, respectively. Finally, it should also be noted that the word sensitivity has been used incorrectly in place of LOD [64].

[C&D-N12]. Experimental design. There are several methods to estimate the limits from simple to complex approaches such as signal-to-noise ratio, standard deviation of blank samples, calibration curve (weighted or not) and pre-established area RSD values [65]. Presenting or discussing pros and cons of the different procedures developed for estimating LOD values are outside the aims of this manuscript. Anyway, in all methods some assumptions and simplifications are applied that are not always acceptable. This fact can significantly influence the estimated values. Additionally, it must be highlighted that the same LOD estimation approach is not automatically usable for all the analytical techniques due to differences in the way that analytical techniques provide instrument signals. Therefore, the LOD estimates obtained by different methodologies are not strictly comparable to each other and they can vary significantly even for the same analytical data [65]. This is the reason that MV guidelines often leave the analyst free to select the LOD acceptance criteria. Two recommendations relating to LOD are: i) The exact procedure for determination of LOD must be clearly stated in the document. If the method of estimation was not visibly indicated, usually results are not valid to be compared; ii) Estimated value for LOD, obtained by theoretical calculation, should be checked to get reliable values. Therefore, it is required the verification of estimate values by the analysis of independent samples around the LOD.

3.2.8. Robustness/Ruggedness

The consistency of an analytical method is addressed to the capacity of remain unaffected when different experimental conditions are deliberately applied so that the results obtained are completely reliable. Experimental conditions influencing the results of analyses are named critical and they should be evaluated and indicated in the validation report [66]. In order to decrease the quantity of tests required to evaluate this validation parameter a Plackett-Burman design with two levels per variable is suggested to be performed [67]. This approach is very efficient when only the main effect of the different factors is evaluated rather than to assess the value of each particular effect.

[C&D-N13]. Terminology. Although robustness and ruggedness have been frequently used interchangeably, they refer to different characteristics and a distinction between them must be made [68]. Some controversy has been reported in the literature because the term robustness was first defined by Youden and Steiner for collaborative studies among different laboratories [69]. Therefore, ruggedness test can be considered as a precision study as a manner to check the transferability of the analytical method. Considering that reproducibility term has been agreed as alternative precision designation for validation purpose, thus it is recommended that ruggedness term should not be applied. On the contrary, robustness term was proposed more recently to measure the capacity of an analytical method to indicate its insensitivity against changes in the normal test conditions at single laboratory level [70]. Although there is a lack of uniformity and certainly a degree of confusion in the analytical literature, there are some factors useful to discriminate between them. Firstly, the test conditions varied (internal/external). Secondly, at which laboratory level (intralaboratory/interlaboratory). Thirdly, the stage when the study should be carried out. Ruggedness (reproducibility) test by interlaboratory studies must be performed at the late stage of MV. On the other hand, robustness test has been planned sometimes at the end of method development and therefore not considered strictly as a performance parameter. Alternatively, performing the test at the end of MV is senseless in avoiding waste of resources thinking in the option that a method is found not to be robust. Therefore, robustness study should be carried out at the start of MV once the method has been optimized, at least to some extent.

4. Evaluation of controversies and discrepancies among MV guidelines

4.1. Overall evaluation of performance parameters for MV

The frequency of the validation parameters included in the MV guidelines were displayed in the Figure 1. These results revealed the high variability in the prevalence of each statistical validation parameter. The performance parameter most frequently included was precision (97%). Following, limit of detection (92%) and selectivity/specificity (89%). Later, calibration/linearity (84%). Accuracy and trueness terms were both used, but the first one was mostly preferred (76% versus 43%). Robustness/ruggedness has a medium/low prevalence (65%). Finally, for many analysts, the value of absolute recovery is not important because it was the performance parameter with lowest presence in the MV guidelines. However, the percentage increases intensely if both concepts (absolute and apparent) of recovery term are merged.

4.2. Particular evaluation of performance parameters for MV

Table 2 summarizes the discrepant information among MV guidelines. Following, the results of each performance parameter are individually evaluated considering, in each case, only the documents including the selected parameter.

4.2.1. Selectivity/Specificity

Different options were used to describe the ability of a method to determine an analyte without interferences from other components. Firstly, many MV guidelines included both terms but selectivity was designed as a preferred term (27%). Secondly, the use of each term alone was very similar for specificity (21%) and selectivity (18%). Another option reported was to use both terms together, as equivalent (21%) or as different (9%) terms. Following, one document included both terms but designating specificity as a preferred term (3%). Finally, it is important to highlight that in three MV documents the general term interference was used to evaluate this performance parameter.

4.2.2. Calibration/Linearity/ Response Function

The preferred terminology for this performance parameter was to use both terms (calibration and linearity) together (48%). Other options reported were to use the single term calibration (29%) or linearity (23%).

In addition, MV guidelines include some general recommendations for preparing the calibration curve:

- Using the same matrix in which the method will be applied later because there are often interactions with matrix components.
- Applying the internal standard, mainly for chromatographic methods, as a way to improve the results obtained.
- A minimum of five-six calibration levels, sometimes suggesting a blank sample (matrix without analyte and internal standard), and a zero sample (matrix without analyte but with internal standard).
- Some discussions still remain concerning the selection of these levels as well as their equidistant or non-equidistant separation.
- Similarly, the number of replicate measurement is widely variable among MV guidelines.
- Unfortunately, only a few documents suggest to study the calibration curve in different series or days (at least three) as a way to evaluate the stability or variability of the instrument response.

MV guidelines were evaluated according to the relationship between concentration and instrument signal. Around 73% of documents included the possibility that the relationship cannot be linear, mainly quadratic. Therefore, about 27% of documents limited the goodness-of- fit to simplest linear model. Relating to the selection of calibration model, that means OLS versus WLS, this decision is critical to avoid biasing the regression line in favour of the calibration standards at high concentration. However, only about 45% of MV guidelines, mainly for biological samples determinations, suggested the use of WLS model and weighting factor (usually 1/x or $1/x^2$). That means WLS model was not included in the majority of documents. In addition, MV guidelines included different procedures to evaluate the goodness-of-fit of the selected calibration model, although the values of r and/or R^2 were selected in 61% of documents. Anyway, around half of these MV guidelines criticize the use of r and/or R^2 as a good indicator to evaluate the goodness-of-fit. On the other hand, %RE was suggested in only 9 out of 33 documents (27%) where acceptance criteria were included, being recommended in 4 out of 8 (50%) analytical MV guidelines for biological matrices.

4.2.3. Accuracy

The evaluation of selected MV guidelines clearly showed that there is no consensus at all on the definition of accuracy. On the one hand, 57% of documents that used this term refer to a single performance parameter. On the other hand, in 43% of documents accuracy was considered as a dual parameter concept serving to define the total analytical error.

Analogous results of lack of agreement were obtained in the evaluation of accuracy (or trueness) and precision by using combined or separate experiments. Exactly, 57% versus

43% of MV guidelines suggested to evaluate them by using single or combined experiments, respectively.

4.2.4. Precision

Though official guidelines suggested the precision levels were widely variable. In the majority of documents, the three typical levels (repeatability, intermediate precision and reproducibility) were reported (44%). Later, only repeatability and intermediate precision were suggested in many MV guidelines (36%). Other minor options reported were to evaluate only repeatability and reproducibility (8%) or repeatability alone (6%). Unfortunately, there are only two documents (6%) that include the four types of precision, which means adding the instrument precision to the three typical levels of method precision. Other subject of interest is the terminology used to define the variability of results when the experimental conditions are varied at single laboratory. The term intermediate precision was used in 42% of documents. Different alternative terms were selected such as within-lab reproducibility, intralaboratory reproducibility, within-run precision, internal precision, run to run precision. Exceptionally, two documents used the term ruggedness for this kind of intermediate precision.

4.2.5. Trueness

It is important to note that this performance parameter is particularly controversial and a typical case of mistaken terminology used in several MV guidelines. Firstly, the terms accuracy and trueness are used as synonymous. Secondly, the trueness (or accuracy) of an analytical method was quantitatively expressed using different terms such as bias, relative bias or recovery. The evaluation of selected MV guidelines showed that the terms used were recovery (41%), bias and recovery (34%) and bias (25%). Surprisingly, five documents had no information at all about systematic error nomenclature.

On the other hand, it was previously commented the significance that the term apparent recovery should be used unequivocally instead of recovery to express the ratio of the concentration found versus the reference value. Probably due to nomenclature simplification but, considering that many documents (75%) include recovery term in the text of MV guidelines, it is difficult to understand that only two documents such as Eurachem [19] and NMKL [29] included apparent recovery as the correct terminology. Additionally, IUPAC guideline [25] for single laboratories used the alternative terms of surrogate or marginal recovery.

4.2.6. Recovery

Significant confusion of the recovery parameter has been observed in the documents. Different validation guidelines (19%) from the total selected, mainly for BMV, refer to recovery from the sample preparation point of view and the term is mostly used as a parameter concerning extraction efficiency. In fact, some guidelines such as ISO 12787 [24] and USFDA-CDER-BMV [39] specified that recovery is related to extraction efficiency. However, there are some exceptions and recovery term was not mentioned in EMA guideline [17]. The organisation argues that recovery is an issue to be investigated during the analytical method development and as such is not considered to be included in the MV guideline. On the other hand, although recovery is described in some documents as a particular performance parameter, really it was previously explained that recovery term is used wrongly as a measure of accuracy/trueness. In any case, interpretation of recovery

from extraction or spiking point of view can be considered as a significant subject from MV guidelines evaluated.

4.2.7. Limit of detection

This is a performance parameter with serious differences in terminology, the experimental procedure and the method of calculation. Firstly, LOD term was used in the majority of MV guidelines (50%). Alternative terms were detection limit (24%), method detection limit (9%), low limit of quantitation (9%) and *CCβ* (6%). Secondly, relating to the methodology for calculation, there are many MV guidelines where this information is missing (41%). Alternatively, more than one method of calculation was reported in 32% of documents while the use of blank samples was suggested in 21% of documents. Lastly, the method of calibration curve and the signal to noise ratio was used exceptionally one time each (3%). On the other hand, it was previously explained that the only way to get reliable LOD values is by verification of the theoretical values obtained. Unfortunately, the recommendation for checking the theoretic results was only incorporated in 5 out of 34 documents (15%) where this parameter was assessed.

4.2.8. Robustness/Ruggedness

Both terms are used to express the consistency of an analytical method when different experimental conditions are intentionally applied. Ruggedness is the term preferred in the majority of MV guidelines (42%). It is important to highlight that this is a very controversial subject because ruggedness was a term selected to check the variability of results among different laboratories and the majority of documents evaluated are relating to single laboratory validation. Alternatively, robustness/ruggedness together have been used in 33% of documents. However, the utilization of robustness, which can be considered as the correct term, was suggested only in 25% of documents.

5. Suggestions by the authors

From this review manuscript, the terms that should be used for analytical MV are:

- Selectivity, as a measure of interference in the process.
- Response function and goodness-of-fit, when choosing the calibration model.
- Accuracy, as a two component parameter formed by precision and trueness.
- Repeatability, intermediate precision and reproducibility, as the terms to define the precision or the method random error.
- Trueness, as the general characteristic to measure the systematic error. In addition, bias or apparent recovery should be used unambiguously when referring to the measurement of systematic error.
- Recovery, should be limited when a study is focused in the concentration or extraction stage.
- Limit of detection, or detection limit, as a form to define statistically the confidence of measurement at low concentrations.
- Robustness, as the consistency of an analytical method at single laboratory level.

Some suggestions for other controversial subjects corresponding to experimental procedure and acceptance criteria of analytical MV are:

- Instrument precision should be complementary firstly evaluated to the three typical method precision levels.

- Calibration curve should be selected including the options of a non-linear and WLS models.
- Goodness-of-fit for calibration model should be never based on r and or R^2 values. The parameter to take into account for evaluation should be % RE of back calculated concentrations.
- Accuracy study should be carried out by combined experiments of precision and trueness using different samples from calibration process.
- Methodology used to evaluate theoretical LOD values should always be reported. Additionally, these values should be verified experimentally at laboratory level.

6. Conclusions

When selecting an analytical method to be used at the laboratory, its validity depends on the particular MV guideline selected because there are many options which can differ in terminology, experimental procedure and acceptance criteria. The main problem among MV guidelines is relating to the terminology used in the different analytical fields. Unfortunately, the diverse performance parameters are not always clearly defined in order to avoid suspicious MV procedures. Therefore, a consensus on a common terminology for validation is required. Similarly, agreement in the experimental procedure and acceptance criteria is also a requisite to try to harmonize method validation practice in all the analytical fields.

Acknowledgements

The authors wish to express their gratitude to the Ministry of Innovation, Science and Universities (Project number CTM2017-83870-R) for the financial support. Thanks to Coordination for the Improvement of Higher Education Personnel (CAPES) for the international exchange fellowship of Carolina Ibelli Bianco (PDSE-Process 88881.189479/2018-01.

References

[1] G.A. Shabir, Validation of high-performance liquid chromatography methods for pharmaceutical analysis - Understanding the differences and similarities between validation requirements of the US Food and Drug Administration, the US Pharmacopeia and the International Conference on Harmonization, J. Chromatogr. A. 987 (2003) 57–66. https://doi.org/10.1016/s0021-9673(02)01536-4.

[2] E. Rozet, A. Ceccato, C. Hubert, E. Ziemons, R. Oprean, S. Rudaz, B. Boulanger, P. Hubert, Analysis of recent pharmaceutical regulatory documents on analytical method validation, J. Chromatogr. A. 1158 (2007) 111–125. https://doi.org/10.1016/j.chroma.2007.03.111.

[3] S. Chandran, R.S.P. Singh, Comparison of various international guidelines for analytical method validation, Pharmazie. 62 (2007) 4–14. https://doi.org/10.1691/ph.2007.1.5064.

[4] D. Stöckl, H. D'Hondt, L.M. Thienpont, Method validation across the disciplines - Critical investigation of major validation criteria and associated experimental protocols, J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 877 (2009) 2180–2190. https://doi.org/10.1016/j.jchromb.2008.12.056.

[5] S. Kollipara, G. Bende, N. Agarwal, B. Varshney, J. Paliwal, International guidelines for bioanalytical method validation: A comparison and discussion on current scenario, Chromatographia. 73 (2011) 201–217. https://doi.org/10.1007/s10337-010-1869-2.

[6] N. Kadian, K.S.R. Raju, M. Rashid, M.Y. Malik, I. Taneja, M. Wahajuddin, Comparative assessment of bioanalytical method validation guidelines for pharmaceutical industry, J. Pharm. Biomed. Anal. 126 (2016) 83–97. https://doi.org/10.1016/j.jpba.2016.03.052.

[7] American Academy Forensic Sciences (AAFS); Academic Standard Board (ASB); Approved American National Standard (ANSI). Standard 036, First Edition. Standard Practices for Method Validation in Forensic Toxicology (2019). http://www.asbstandardsboard.org/wp-content/uploads/2019/11/036_Std_e1.pdf.

[8] Brazilian Sanitary Surveillance Agency (ANVISA). Resolution - RE n. 899, of May 29, 2003. Guide for validation of analytical and bioanalytical methods. Official Diary of the Union, June 02 (2003). http://portal.anvisa.gov.br/legislacao/?inheritRedirect=true#/visualizar/27342.

[9] Association of Analytical Communities (AOAC). AOAC Guidelines for Single Laboratory Validation of Chemical Methods for Dietary Supplements and Botanicals. (2002). http://members.aoac.org/aoac_prod_imis/AOAC_Docs/StandardsDevelopment/SLV_Guidelines_Dietary_Supplements.pdf.

[10] Australian Pesticides & Veterinary Medicines Authority (APVMA). Guidelines for the validation of analytical methods for active constituent, agricultural and veterinary chemical products (2004). https://apvma.gov.au/sites/default/files/docs/guideline-69-analytical-methods.pdf.

[11] American Society for Testing and Materials (ASTM). Designation: E2857-11. Standard Guide for Validating Analytical Methods. ASTM International, West Conshohocken, PA. (2011). https://doi.org/10.1520/E2857.

[12] Commission Decision of European Union. Council Directive 96/23/EC. The performance of analytical methods and the interpretation of results. Off. J. Eur. Communities. (2002). https://op.europa.eu/en/publication-detail/-/publication/ed928116-a955-4a84-b10a-cf7a82bad858/language-en.

[13] The European Committee for Standardization (CEN). Guide 13. Validation of environmental test methods. (2008). https://boss.cen.eu/ref/CEN_13.pdf.

[14] Collaborative International Pesticides Analytical Council (CIPAC). Guidelines on method validation to be performed in support of analytical methods for agrochemical formulations. (2003). https://www.cipac.org/images/pdf/validat.pdf.

[15] S. Bratinova, B. Raffael, C. Simoneau, Community and National Reference Laboratories (CRL and NRL); Food Contact Materials (FCM). EUR 24105 EN. Guidelines for performance criteria and validation procedures of analytical methods used in controls of food contact materials. (2009). https://doi.org/10.2788/49046.

[16] EDES, a COLEACP programme (Europe-Africa-Caribbean-Pacific Liaison Committee). Handbook 8.5. Management of Laboratories: Method validation. (2013). https://eservices.coleacp.org/en/e-bibliotheque/85-method-validation-0.

[17] European Medicines Agency (EMA), Committee for Medicinal Products for Human Use (CHMP), Efficacy Working Party (EWP). EMEA/CHMP/EWP/192217/2009 Rev. 1 Corr. 2. Guideline on bioanalytical method validation. Eur. Med. Agency. (2011). https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-bioanalytical-method-validation_en.pdf.

[18] European Network of Forensic Science Institutes (ENFSI). Guidelines for the single laboratory Validation of Instrumental and Human Based Methods in Forensic Science. Version 2.0. (2014). http://enfsi.eu/wp-content/uploads/2017/06/Guidelines-for-the-single-laboratory-Validation-of-Instrumental-and-Human-Based-Methods-in-Forensic-Sciene_2014-version-2.0.pdf.

[19] Magnusson, B.; Örnemark, U. (eds.). Eurachem Guide. The Fitness for Purpose of Analytical Methods: A Laboratory Guide to Method Validation and Related Topics. Second edition. ISBN 978-91-87461-59-0. (2014). https://www.eurachem.org/index.php/publications/guides/mv.

[20] Food & Agriculture Organization/ International Atomic Energy Agency (FAO/IAEA). Validation of Analytical Methods for Food Control. (1998). http://www.fao.org/3/a-w8420e.pdf.

[21] F. Peters, M. Hartung, M. Herbold, G. Schmitt, T. Daldrup, F. Mußhoff, Society of Toxicological & Forensic Chemistry (GTFCh). Guidelines for Quality Control in Forensic-Toxicological Analyses. (2009). https://pdfs.semanticscholar.org/7c2a/2218f37b98ef7f8d684c8b6a497ebf0026ca.pdf?_ga=2.66711790.1511258561.1583167409-58668952.1570826285.

[22] International Conference on Harmonisation (ICH) of technical requirements for registration of pharmaceuticals for human use. ICH harmonised tripartite guideline. Validation of analytical procedures: Text and Methodology Q2(R1). (2005). https://database.ich.org/sites/default/files/Q2_R1__Guideline.pdf.

[23] Irish National Accreditation Board (INAB). PS 15. Guide to Method Validation for Quantitative Analysis in Chemical Testing Laboratories. (2019). https://www.inab.ie/Documents-Forms/Policy/Guide-to-Method-Validation-for-Quantitative-Analysis-in-Chemical-Testing-Laboratories-17025-PDF-36-Pages-349KB-.pdf.

[24] International Organization for Standardization (ISO). ISO 12787:2011 (E). Cosmetics - Analytical methods - Validation criteria for analytical results using chromatographic techniques. (2011). https://www.iso.org/standard/51709.html.

[25] M. Thompson, S.L.R. Ellison, R. Wood, Harmonized guidelines for single-laboratory validation of methods of analysis (IUPAC Technical Report), Pure Appl. Chem. 74 (2002) 835–855. https://doi.org/10.1351/pac200274050835.

[26] Ministry of Health, Labour and Welfare (MHLW) Japan. Guideline on bioanalytical method validation in pharmaceutical development. (2013). https://www.nihs.go.jp/drug/BMV/250913_BMV-GL_E.pdf.

[27] National Association of Testing Authorities (NATA) Australia. General Accreditation Guidance – Validation and verification of quantitative and qualitative test methods (2018).https://www.nata.com.au/phocadownload/gen-accreditation-guidance/Validation-and-Verification-of-Quantitative-and-Qualitative-Test-Methods.pdf.

[28] National Environmental Laboratory Accreditation Institute (NELAC-TNI). Volume 1, Module 4. Chemical testing. (2016). https://nelac-institute.org/index.php.

[29] Nordic Committee on Food Analysis (NMKL). Procedure No. 4. Validation of chemical analytical methods. (2009). https://www.nmkl.org/index.php/en/publications/item/validering-av-kjemiske-analysemetoder-nmkl-prosedyre-nr-4-2009.

[30] Nordic Validation International (NordVal). Protocol No. 2. Guide in validation of alternative proprietary chemical methods. (2010). http://members.aoac.org/aoac_prod_imis/AOAC_Docs/ISPAM/3.9NordValprotocolproprietarychemicalanalysis.pdf.

[31] Organisation for Economic Co-operation and Development (OECD). ENV/JM/MONO(2014)20. Series on Testing and Assessment No. 204 and Series on Biocides No. 9. Guidance document for single laboratory validation of quantitative analytical methods – Guidance use. (2014). http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2014)20&doclanguage=en.

[32] International Organization of Vine & Wine (OIV). Resolution OENO 8/2005. Harmonised guidelines for single-laboratory validation of methods of analysis (technical report). (2005). http://www.oiv.int/public/medias/780/oeno-8-2005-en.pdf.

[33] European Commission. Safety of the Food Chain Pesticides and Biocides. SANTE/11813/2017. Guidance document on analytical quality control and method validation procedures for pesticides residues analysis in food and feed. (2017). https://ec.europa.eu/food/sites/food/files/plant/docs/pesticides_mrl_guidelines_wrkdoc_2017-11813.pdf.

[34] P. Hubert, J.J. Nguyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P.A. Compagnon, W. Dewé, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard,
C. Nivet, L. Valat, E. Rozet, Harmonization of strategies for the validation of quantitative analytical procedures. A SFSTP proposal - Part II, J. Pharm. Biomed. Anal. 45 (2007) 70–81. https://doi.org/10.1016/j.jpba.2007.06.013.

[35] Scientific Working Group for Forensic Toxicology (SWGTOX). Report from the Scientific Working Group for Forensic Toxicology. Standard Practices for Method Validation in Forensic Toxicology. J. Anal. Toxicol. 37 (2013) 452–474. https://doi.org/10.1093/jat/bkt054.

[36] United States Environmental Protection Agency (US EPA). Guidance for Methods Development and Methods Validation for the Resource Conservation and Recovery Act (RCRA) Program. (1992). https://www.epa.gov/sites/production/files/2015-10/documents/methdev.pdf.

[37] U.S. Environmental Protection Agency (EPA) Forum on Environmental Measurements (FEM). Validation and Peer Review of U.S. Environmental Protection Agency Chemical Methods of Analysis. (2016). https://www.epa.gov/sites/production/files/2015-01/documents/chemmethod_validity_guide.pdf.

[38] United States Food & Drug Administration (US FDA); Centre for Drug Evaluation & Research (CDER). Validation of chromatographic methods. (1994). https://www.fda.gov/media/75643/download.

[39] U.S. Department of Health and Human Services Food and Drug Administration (US FDA); Center for Drug Evaluation and Research (CDER); Center for Veterinary Medicine (CVM). Bioanalytical Method Validation: Guidance for Industry. (2018). https://www.fda.gov/files/drugs/published/Bioanalytical-Method-Validation-Guidance-for-Industry.pdf.

[40] U.S. Food and Drug Administration/Foods and Veterinary Medicine (FDA/FVM). Guidelines for the Validation of Chemical Methods for the FDA Foods Program. 3rd Edition. (2019). https://www.fda.gov/media/81810/download.

[41] United States Pharmacopeial (USP). <1225>. Validation of Compendial Procedures. (2016). http://www.ofnisystems.com/wp-content/uploads/2013/12/USP36_1225.pdf.

[42] World Health Organization (WHO). Guidelines on validation - Appendix 4 - Analytical method validation. (2018). https://www.who.int/medicines/areas/quality_safety/quality_assurance/28092018Guideline_Validation_AnalyticalMethodValidation-Appendix4_QAS16-671.pdf.

[43] Joint Committee for Guides in Metrology (JCGM). International vocabulary of metrology - Basic and general concepts and associated terms (VIM), 3rd edn. JCGM 200:2012. (2012). https://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf.

[44] P. Hubert, P. Chiap, J. Crommen, B. Boulanger, E. Chapuzet, N. Mercier, S. Bervoas-Martin, P. Chevalier, D. Grandjean, P. Lagorce, M. Lallier, M.C. Laparra, M. Laurentie, J.C. Nivet, The SFSTP guide on the validation of chromatographic methods for drug bioanalysis: From the Washington Conference to the laboratory, Anal. Chim. Acta. 391 (1999) 135–148. https://doi.org/10.1016/S0003-2670(99)00106-3.

[45] C. Hartmann, J. Smeyers-Verbeke, D.L. Massart, R.D. McDowall, Validation of bioanalytical chromatographic methods, J. Pharm. Biomed. Anal. 17 (1998) 193–218. https://doi.org/10.1016/S0731-7085(97)00198-2.

[46] M. Rambla-Alegre, J. Esteve-Romero, S. Carda-Broch, Is it really necessary to validate an analytical method or not? That is the question, J. Chromatogr. A. 1232 (2012) 101–109. https://doi.org/10.1016/j.chroma.2011.10.050.

[47] J. Vessman, Selectivity or specificity? Validation of analytical methods from the perspective of an analytical chemist in the pharmaceutical industry, J. Pharm. Biomed. Anal. 14 (1996) 867–869. https://doi.org/10.1016/0731-7085(95)01679-1.

[48] H.Y. Aboul-Enein, Selectivity versus specificity in chromatographic analytical methods, Accredit. Qual. Assur. 5 (2000) 180–181. https://doi.org/10.1007/s007690050440.

[49] J. Vessman, R.I. Stefan, J.F. Van Staden, K. Danzer, W. Lindner, D.T. Burns, A. Fajgelj, H. Müller, Selectivity in analytical chemistry, Rev. Roum. Chim. 73 (2001) 1381–1386. https://doi.org/10.1351/pac200173081381.

[50] P. Araujo, Key aspects of analytical method validation and linearity evaluation, J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 877 (2009) 2224–2234. https://doi.org/10.1016/j.jchromb.2008.09.030.

[51] J.M. Jurado, A. Alcázar, R. Muñiz-Valencia, S.G. Ceballos-Magaña, F. Raposo, Some practical considerations for linearity assessment of calibration curves as function of concentration levels according to the fitness-for-purpose approach, Talanta. 172 (2017) 221–229. https://doi.org/10.1016/j.talanta.2017.05.049.

[52] A.M. Almeida, M.M. Castel-Branco, A.C. Falcão, Linear regression for calibration lines revisited: Weighting schemes for bioanalytical methods, J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 774 (2002) 215–222. https://doi.org/10.1016/S1570-0232(02)00244-1.

[53] F. Raposo, Evaluation of analytical calibration based on least-squares linear regression for instrumental techniques: A tutorial review, TrAC - Trends Anal. Chem. 77 (2016) 167–185. https://doi.org/10.1016/j.trac.2015.12.006.

[54] International Organization for Standardization (ISO). ISO 5725. Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions, (1994) 1–17. https://www.iso.org/obp/ui/#iso:std:iso:5725:-1:ed-1:v1:en.

[55] H.T. Karnes, C. March, Precision, accuracy, and data acceptance criteria in biopharmaceutical analysis, Pharm. Res. An Off. J. Am. Assoc. Pharm. Sci. 10 (1993) 1420–1426. https://doi.org/10.1023/A:1018958805795.

[56] A.R. Buick, M. V. Doig, S.C. Jeal, G.S. Land, R.D. McDowall, Method validation in the bioanalytical laboratory, J. Pharm. Biomed. Anal. 8 (1990) 629–637. https://doi.org/10.1016/0731-7085(90)80093-5.

[57] M. Thompson, Precision in chemical analysis: A critical survey of uses and abuses, Anal. Methods. 4 (2012) 1598–1611. https://doi.org/10.1039/c2ay25083g.

[58] R. Boqué, A. Maroto, J. Riu, F.X. Rius, Validation of Analytical Methods, Grasas y Aceites. 53 (2002) 128–143. https://doi.org/10.3989/gya.2002.v53.i1.295.

[59] T.P.J. Linsinger, Use of recovery and bias information in analytical chemistry and estimation of its uncertainty contribution, TrAC - Trends Anal. Chem. 27 (2008) 916– 923. https://doi.org/10.1016/j.trac.2008.08.013.

[60] D.T. Burns, K. Danzer, A. Townshend, Use of the terms “recovery” and “apparent recovery” in analytical procedures (IUPAC Recommendations 2002), Pure Appl. Chem. 74 (2002) 2201–2205. https://doi.org/10.1351/pac200274112201.

[61] L.A. Currie, Nomenclature in evaluation of analytical methods including detection and quantification capabilities (IUPAC Recommendations 1995), Anal. Chim. Acta. 391 (1999) 105–126. https://doi.org/10.1016/S0003-2670(99)00104-X.

[62] N.M. Faber, The limit of detection is not the analyte level for deciding between “detected” and “not detected,” Accredit. Qual. Assur. 13 (2008) 277–278. https://doi.org/10.1007/s00769-007-0351-9.

[63] M. Belter, A. Sajnóg, D. Barałkiewicz, Over a century of detection and quantification capabilities in analytical chemistry - Historical overview and trends, Talanta. 129 (2014) 606–616. https://doi.org/10.1016/j.talanta.2014.05.018.

[64] J.L. Rudy, Differentiating between Sensitivity and Limit of Detection, Clin. Chem. 35 (1989) 509. https://doi.org/10.1093/clinchem/35.3.509.

[65] H. Evard, A. Kruve, I. Leito, Tutorial on estimating the limit of detection using LC-MS analysis, part I: Theoretical review, Anal. Chim. Acta. 942 (2016) 23–39. https://doi.org/10.1016/j.aca.2016.08.043.

[66] Y. Vander Heyden, A. Nijhuis, J. Smeyers-Verbeke, B.G.M. Vandeginste, D.L. Massart, Guidance for robustness/ruggedness tests in method validation, J. Pharm. Biomed. Anal. 24 (2001) 723–753. https://doi.org/10.1016/S0731-7085(00)00529-X.

[67] N.A. Epshtein, Validation of HPLC techniques for pharmaceutical analysis, Pharm. Chem. J. 38 (2004) 212-228. https://doi.org/10.1023/B:PHAC.0000038422.27193.6c.

[68] D.T. Burns, K. Danzer, A. Townshend, A tutorial discussion of the use of the terms “Robust” and “Rugged” and the associated characteristics of “Robustness” and “Ruggedness” as used in descriptions of analytical procedures, J. Assoc. Public Anal. 37 (2009) 40–60. http://www.apajournal.org.uk/2009_0040_0060.pdf.

[69] W.J. Youden, E.H. Steiner, Statistical manual of the Association of Official Analytical Chemists,1975. http://agris.fao.org/agris-search/search.do?recordID=XF2015020336.

[70] Y. Vander Heyden, J. Saevels, E. Roets, J. Hoogmartens, D. Decolin, M.G. Quaglia,
W. Van Den Bossche, R. Leemans, O. Smeets, F. Van De Vaart, B. Mason, G.C. Taylor, W. Underberg, A. Bult, P. Chiap, J. Crommen, J. De Beer, S.H. Hansen, D.L. Massart, Interlaboratory studies on two high-performance liquid chromatographic assays for tylosin (tartrate), J. Chromatogr. A. 830 (1999) 3–28. https://doi.org/10.1016/S0021-9673(98)00840-1.

Figure Captions

Figure 1.- Frequency of validation parameters included in MV guidelines.

Table 1. Summary of the analytical method validation guidelines evaluated.

GUIDE	ACRONYM	ORGANIZATION NAME	NATIONAL or INTERNATIONAL	MATRICES	AREA or DISCIPLINE	YEAR	REFERENCE
1	AAFS-ASB	American Academy Forensic Sciences Academic Standard Board	National	Biological	forensic	2019	[7]
2	ANVISA	Brazilian Sanitary Surveillance Agency	National	Analytical	pharmaceutical	2003	[8]
3	ANVISA	Brazilian Sanitary Surveillance Agency	National	Biological	drugs	2003	[8]
4	AOAC	Association of Analytical Communities	International	Analytical	foods	2002	[9]
5	APVMA	Australian Pesticides & Veterinary Medicines Authority	National	Analytical	active constituents, agricultural and veterinary chemical products	2004	[10]
6	ASTM	American Society for Testing and Materials	International	Analytical	metals, ores materials	2011	[11]
7	CD 96/23/EC	Commission Decision of European Union	International	Analytical	residues in products of animal origin	2002	[12]
8	CEN	The European Committee for Standardization	International	Analytical	environmental samples	2008	[13]
9	CIPAC	Collaborative International Pesticides Analytical Council	International	Analytical	agrochemical formulations	2003	[14]
10	CRL-NRL-FCM	Community and National Reference Laboratories Food Contact Materials	International	Analytical	food contact materials	2009	[15]
11	EDES	Europe and Africa, Caribbean and Pacific countries	International	Analytical	food and feedstuffs	2013	[16]
12	EMA	The European Medicines Agency	International	Biological	drugs	2011	[17]
13	ENFSI	The European Network of Forensic Science Institutes	International	Biological	forensic	2014	[18]
14	EURACHEM	Eurachem	International	Analytical	not specified	2014	[19]
15	FAO-IAEA	Food & Agriculture Organization International Atomic Energy Agency	International	Analytical	food	1998	[20]
16	GTFCh	The Society of Toxicological & Forensic	International	Biological	forensic	2009	[21]

GUIDE	ACRONYM	ORGANIZATION NAME	NATIONAL or INTERNATIONAL	MATRICES	AREA or DISCIPLINE	YEAR	REFERENCE
		Chemistry					
17	ICH	The International Council for Harmonization of Technical Requirements for Pharmaceuticals	International	Analytical	pharmaceutical	2005	[22]
18	INAB	The Irish National Accreditation Board	National	Analytical	chemical analysis	2019	[23]
19	ISO 12787	The International Organization for Standardization	International	Analytical	cosmetics	2011	[24]
20	IUPAC	The International Union of Pure & Applied Chemistry	International	Analytical	not specified	2002	[25]
21	MHLW	The Ministry of Health, Labour and Welfare-Japan	National	Biological	drugs	2013	[26]
22	NATA	The National Association of Testing Authorities - Australia	National	Analytical	not specified	2018	[27]
23	NELAC-TNI	The National Environmental Laboratory Accreditation Institute	National	Analytical	environmental samples	2016	[28]
24	NMKL	The Nordic Committee on Food Analysis	International	Analytical	food, drinking water or animal feed	2009	[29]
25	NORD-VAL	The Nordic Validation International	International	Analytical	chemical methods (test kits)	2010	[30]
26	OECD	The Organization of Economic Co-Operation & Development	International	Analytical	biocides	2014	[31]
27	OIV	The International Organization of Vine & Wine	International	Analytical	wine	2005	[32]
28	SANTE	The Directorate-General for Health and Food safety	International	Analytical	pesticide residues and analysis in food and feed	2017	[33]
29	SFSTP	The French Society of Pharmaceutical Sciences & Techniques	National	Analytical	pharmaceutical	2007	[34]
30	SWGTOX	The Scientific Working Group for Forensic Toxicology	International	Biological	forensic	2013	[35]
31	USEPA	The United States Environmental Protection Agency	National	Analytical	environmental samples	1992	[36]
32	USEPA-FEM		National	Analytical	chemical methods	2016	[37]

GUIDE	ACRONYM	ORGANIZATION NAME	NATIONAL or INTERNATIONAL	MATRICES	AREA or DISCIPLINE	YEAR	REFERENCE
		USEPA-Forum on Environmental Measurements					
33	USFDA-CDER	Centre for Drug Evaluation & Research	National	Analytical and biological	chromatographic test methods	1994	[38]
34	USFDA-CDER-CVM	Centre for Veterinary Medicine	National	Biological	drugs	2018	[39]
35	USFDA-FVM	Foods & Veterinary Medicine Program	National	Analytical	food, feed, cosmetics, and veterinary products	2019	[40]
36	USP	The United States Pharmacopeia	National	Analytical	pharmaceutical	2016	[41]
37	WHO	The World Health Organization	International	Analytical	medicines	2018	[42]

Table 2. Controversies and discrepancies (C&D) among the evaluated analytical method validation guidelines.

Guide	(N1) SEL	(N2) LIN-1	(N3) LIN-2	(N4) LIN-3	(N5) ACC-1	(N6) ACC-2	(N7) PREC-1	(N8) PREC-2	(N9) TRUE	(N10) RECO	(N11) LOD-1	(N12) LOD-2	(N13) ROBU	REF.
1	INTERF.	CAL	2/WLS	YES[1]	X	COMB	1/2	NO/run	NO/bias	X	YES	VAR[5]	X	[7]
2	SEL/SPE	LIN	1/OLS	YES	1	X	1/2/3	IP/run	NO/recov	NO	YES	Cal	ROBU	[8]
3	SPE	CAL/LIN	2/OLS	YES[3]	1	X	1/2	NO/run	NO/recov	YES	NO/DL	None	X	[8]
4	SEL (SPE)	CAL	2/WLS	YES[1]	1	X	1/2/3	IP/labor	NO/recov	NO	NO/determ	Blanks	RUGG	[9]
5	SEL (SPE)	LIN	2/OLS	YES	1	X	1/2/3	IP	NO/recov	NO	YES	SDlowconc	X	[10]
6	SEL	CAL	NO/NO	NO[3]	2	X	1/2/3	IP/labor	NO/bias	X	YES	None	RUGG	[11]
7	SPE	X	X	X	2	X	1/2/3	NO/wlrepr	YES/recov	YES (error)	NO/CCβ	Cal/Blanks	RUGG	[12]
8	X	X	X	X	X	X	1/3	X	X	X	X	X	ROBU	[13]
9	SPE	LIN	2/NO	YES	1	X	1	X	NO/recov	NO	X	X	X	[14]
10	SEL/SPE	CAL/LIN	2/WLS	YES[2/3]	2	X	1/2/3	IP/wlrepr	YES/bias-rec	NO	YES/MDL	VAR	ROBU/RUGG	[15]
11	SEL/SPE	CAL/LIN	NO/WLS	NO	2	X	1/2/3	NO/wlrepr	YES/bias	YES	YES	Blanks	ROBU/RUGG	[16]
12	SEL&SPE	CAL	NO/NO	NO[3]	1	COMB	1/2	NO/run	X	X	NO/LLOQ	None	X	[17]
13	SEL (SPE)	LIN	NO/NO	NO	X	X	1/2	NO/wlrepr	YES/bias	X	YES	None	ROBU/RUGG	[18]
14	SEL	CAL/LIN	NO/NO	NO	2	X	1/2/3	IP	YES/bias-rec[4]	NO	YES	Blanks	ROBU/RUGG	[19]
15	SPE	X	X	X	1	X	1/2/3	NO/wlrepr	NO/recov	NO	YES	None	X	[20]
16	SEL (SPE)	CAL/LIN	2/WLS	NO	2	COMB	1/2/3	IP	YES/bias	YES	YES	SNR/Cal	ROBU/RUGG	[21]
17	SPE	LIN	NO/OLS	YES	1	X	1/2/3	IP	NO/recov	NO	YES	VAR[5]	ROBU	[22]
18	SEL (SPE)	CAL/LIN	2/WLS	YES[1]	2	X	1/2	NO/intlabrepr	YES/bias-rec	YES (error)	NO/DL	Blanks	ROBU/RUGG	[23]
19	SEL&SPE	CAL/LIN	2/WLS	YES	1	X	1/2/3	IP	NO/recov	YES	YES	SNR/Cal	X	[24]
20	SEL	CAL/LIN	2/WLS	YES[1]	X	X	1/2	NO/run	YES/bias-rec	YES (error)	YES/DL	Blanks	RUGG	[25]
21	SEL (SPE)	CAL	2/WLS	NO[3]	1	COMB	1/2	NO/run	X	YES	NO/LLOQ	None	X	[26]
22	SEL (SPE)	CAL/LIN	2/WLS	YES[1]	2	X	0/1/2/3	IP/wl-intr repr	YES/bias-rec	NO	YES	VAR	RUGG	[27]
23	SEL	CAL	2/OLS	YES[3]	X	COMB	X	NO	NO/bias-rec	NO	NO/MDL	None[5]	X	[28]
24	SPE	ST.CURV.	2/NO	YES[1]	X	X	1/2/3	NO/inter repr	YES/recov[4]	NO	YES	Blanks/Cal	RUGG	[29]
25	SPE	X	X	X	X	X	1/2	NO/inter repr	YES/bias-rec	YES (error)	NO/CCβ	Blanks	RUGG	[30]
26	SEL/SPE	CAL/LIN	2/NO	YES	1	X	1	NO	NO/recov	NO	YES	None	X	[31]
27	SEL	CAL/LIN	2/WLS	YES[1]	X	X	1/2	NO/runtorun	YES/bias-rec	YES (error)	YES/DL	None	RUGG	[32]

Guide	(N1) SEL	(N2) LIN-1	(N3) LIN-2	(N4) LIN-3	(N5) ACC-1	(N6) ACC-2	(N7) PREC-1	(N8) PREC-2	(N9) TRUE	(N10) RECO	(N11) LOD-1	(N12) LOD-2	(N13) ROBU	REF.
28	SEL (SPE)	CAL/LIN	2/WLS	NO[3]	2	COMB	1/2	NO/wlrepr	YES/bias-rec	NO	X	X	ROBU	[33]
29	SEL/SPE	CAL/LIN/R.F.	2/WLS	YES[1/3]	2	COMB	1/2	IP	YES/bias-rec	NO	YES	None	X	[34]
30	INTERF.	CAL	2/WLS	YES[1]	X	COMB	1/2	NO/run	NO/bias	X	YES	VAR	X	[35]
31	INTERF.	X	2/NO	NO	1	X	1/3	NO/longtermpr	NO/bias-rec	NO	NO/MDL	None	RUGG	[36]
32	SEL	CAL	2/NO	NO	2	COMB	1/3	NO	YES/bias	X	NO/DL	None[5]	RUGG	[37]
33	SEL/SPE	LIN	2/NO	YES	1	COMB	0/1/2/3	IP/ruggedness	NO/recov	YES (error)	NO/DL	SNR	ROBU	[38]
34	SEL&SPE	CAL	2/NO	NO[3]	1	COMB	1/2	NO/run	X	YES	NO/LLOQ	None	X	[39]
35	SEL (SPE)	CAL/LIN	NO/NO	NO	2	X	1/2/3	IP	YES/bias	YES	YES	None	ROBU/RUGG	[40]
36	SPE (SEL)	LIN	2/WLS	YES	1	COMB	1/2/3	IP/ruggedness	NO/recov	NO	NO/DL	VAR[5]	ROBU	[41]
37	SEL/SPE	CAL/LIN	NO/NO	NO	1	X	1/2/3	IP	X	YES (error)	NO/DL	VAR	ROBU/RUGG	[42]

Explanation about controversies and discrepancies (C&D) nomenclature of Table 2.

N1 (SEL): used terminology related to "selectivity"
SEL: only the term "selectivity" is considered
SPE: only the term "specificity" is considered
SEL(SPE) or *SPE(SEL):* the terms "selectivity" and "specificity" are distinguished and the execution only of what is outside the parentheses is considered
SEL/SPE: the terms "selectivity" and "specificity" are used as synonyms
SEL&SPE: the terms "selectivity" and "specificity" are distinguished and the execution of both are considered
INTERF.: the term "interference" is used.

N2 (LIN-1): used terminology related to "linearity"
LIN: only the term "linearity" is considered
CAL: only the term "calibration" is considered
CAL/LIN: both terms, "calibration" and "linearity", are considered
CAL/LIN/R.F.: three terms are considered - calibration, linearity and response function
ST.CURV.: the term "standard curve" is considered.

N3 (LIN-2): selection of the calibration model
1: linear equation

N8 (PREC-2): used terminology related to "precision" – the guideline considers "intermediate precision"/other related terms
NO or IP: does not consider "intermediate precision" OR considers it
Other related terms: run, labor, wlrepr (within-laboratory reproducibility), intlabrepr (inter-laboratory reproducibility), wl-intrrepr (within-laboratory reproducibility and intra-laboratory reproducibility), inter repr (internal reproducibility), runtorun, longtermpr (long-term precision), ruggedness

N9 (TRUE): used terminology related to "trueness"
YES: the term "trueness" is used
NO: the term "trueness" is not used
Bias: "trueness" is expressed using the term "bias"
Recov: "trueness" is expressed using the term "recovery"
Bias-rec: "trueness" is expressed using the terms "bias and recovery"
Superscript 4: the guide mentions the term "apparent recovery"

N10 (RECO): used terminology related to "recovery"
YES: "recovery" is considered a specific parameter

2: nonlinear equation
OLS: ordinary model
WLS: weighted model
NO: does not specify about the equation's linearity or about the considered model.

N4 (LIN-3): acceptance criteria
YES: use r and/or R^2 as criterion for goodness-of-fit
NO: does not use r and/or R^2 as criterion for goodness-of-fit
Superscript 1: critique using r and/or R^2
Superscript 2: wrong definition of r and/or R^2
Superscript 3: use percentage of relative error as criterion for goodness-of-fit.

N5 (ACC-1): used terminology related to "accuracy"
1: accuracy as an individual parameter as a measure of the systematic error
2: accuracy as a set of parameters (precision and trueness).

N6 (ACC-2): single versus combined experiments
COMB.: accuracy evaluation is carried out in combination with precision experiments

N7 (PREC-1): precision levels
0: precision is associated with "instrument precision"
1: precision is associated with "repeatability"
2: precision is associated with "intermediate precision"
3: precision is associated with "reproducibility".

NO: "recovery" is not considered a specific parameter
YES (error): really is "apparent recovery".

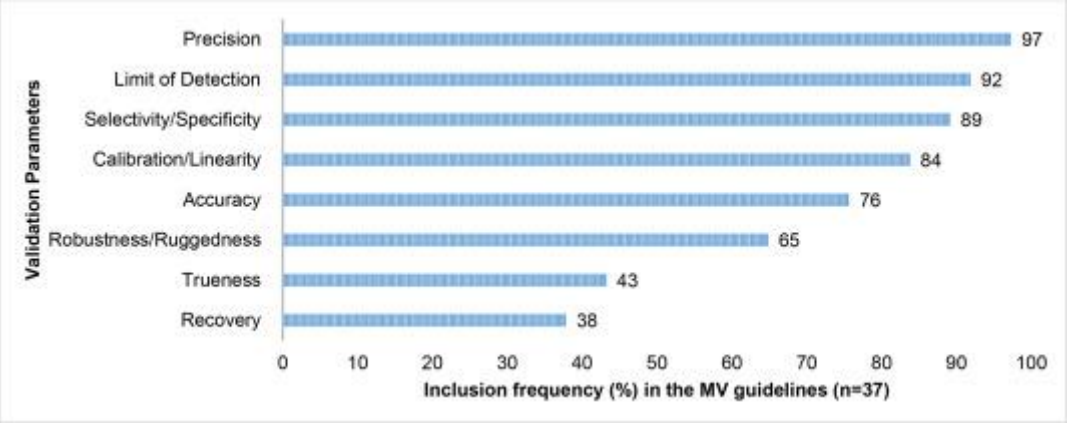
N11 (LOD-1): used terminology related to "limit of detection"
YES: the term "limit of detection" is used
NO: the term "limit of detection" is not used
Alternative designations: DL (detection limit); determ (limit of determination); CCβ (detection capability); MDL (method detection limit); LLOQ (lower limit of quantification).

N12 (LOD-2): suggested method for estimating the "limit of detection"
VAR (various); *None; Blanks; Cal* (calculated); *SDlowconc* (standard deviation - lowest calibration standard); *SNR* (signal-to-noise ratio)
Superscript 5: it is suggested to check LOD experimentally.

N13 (ROBU): used terminology related to "robustness". *ROBU:* only the term "robustness" is considered
RUGG: only the term "ruggedness" is considered
ROBU/RUGG: both terms "robustness and ruggedness" are considered.

X: Information about the parameter is not included in the guideline.

Figura 1


Validation Parameters
Precision
Limit of Detection
Selectivity/Specificity
Calibration/Linearity
Accuracy
Robustness/Ruggedness
Trueness
Recovery
97
92
89
84
76
65
43
38
0
10
20
30
40
50
60
70
80
90
100
Inclusion frequency (%) in the MV guidelines (n=37)